LA-UR-11-0730

*Approved for public release;
distribution is unlimited.*

Title: Extreme Scale Computing and Biosurveillance

Author(s): James P Ahrens 113788 CCS-7
Marcus Daniels 211500 CCS-7

Intended for: Panel V: Global Biosurveillance Information Science and Technology, January 2011

Extreme Scale Computing and Biosurveillance Abstract

“The Key finding of the Panel is that there are compelling needs for exascale computing capability to support the DOE’s missions in energy, national security, fundamental sciences, and the environment. The DOE has the necessary assets to initiate a program that would accelerate the development of such capability to meet it’s own needs and by so doing benefit other national interests. Failure to initiate an exascale program could lead to a loss of U.S. competitiveness in several critical technologies.”

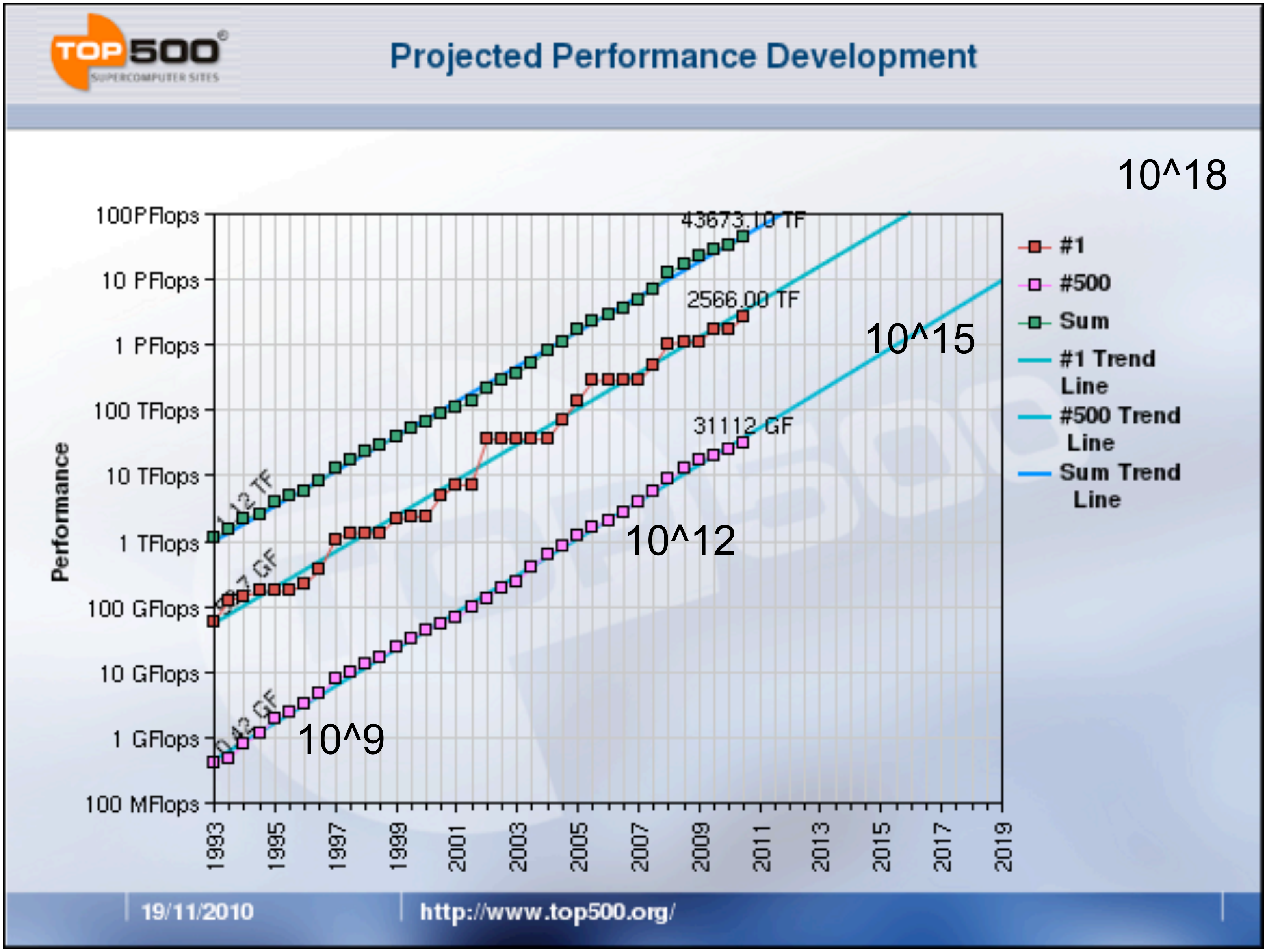
Trivelpiece Panel Report, January 2010

Our goal for the presentation is to add a new perspective to the Biosurveillance community of how Data-Intensive Computing or Super-Computing approach can contribute to Biosurveillance. No matter through high quality images, quantifying the data error for analysis, qualifying visual error for visualization, intelligent sampling designs to provide more information on less data, in-situ & storage-based sampling-based data reduction and more, these will all benefit the Biosurveillance community in future development and enhance U.S. competitiveness in technology.

Extreme Scale Computing and Biosurveillance

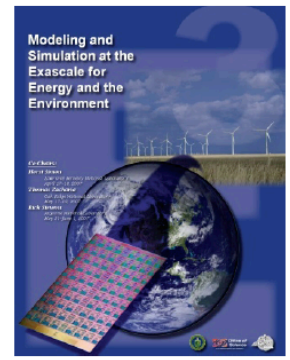
James Ahrens and Marcus Daniels
Los Alamos National Laboratory

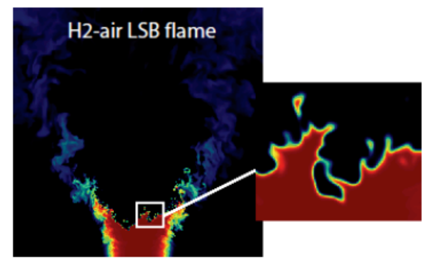
Panel V: Global Biosurveillance Information Science and Technology
January 2011


TOP500
SUPERCOMPUTER SITES
Projected Performance Development
10^18
10^15
10^12
10^9
100PFlops
10 PFlops
1 PFlops
100 TFlops
10 TFlops
1 TFlops
100 GFlops
10 GFlops
1 GFlops
100 MFlops
Performance
43673.10 TF
2566.00 TF
31112 GF
1.12 TF
0.42 GF
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
#1
#500
Sum
#1 Trend Line
#500 Trend Line
Sum Trend Line
19/11/2010
http://www.top500.org/

Exascale Applications and Technology

- **Town Hall Meetings April-June 2007**
- **Scientific Grand Challenges Workshops November 2008 – October 2009**
  - Climate Science (11/08),
  - High Energy Physics (12/08),
  - Nuclear Physics (1/09),
  - Fusion Energy (3/09),
  - Nuclear Energy (5/09) (with NE)
  - Biology (8/09),
  - Material Science and Chemistry (8/09),
  - National Security (10/09) (with NNSA)
- **Cross-cutting workshops**
  - Architecture and Technology (12/09)
  - Architecture, Applied Mathematics and Computer Science (2/10)
- **Meetings with industry (8/09, 11/09)**
- **External Panels**
  - ASCAC Exascale Charge (FACA)
  - Trivelpiece Panel





MISSION IMPERATIVES

"The key finding of the Panel is that there are compelling needs for exascale computing capability to support the DOE's missions in energy, national security, fundamental sciences, and the environment. The DOE has the necessary assets to initiate a program that would accelerate the development of such capability to meet its own needs and by so doing benefit other national interests. Failure to initiate an exascale program could lead to a loss of U. S. competitiveness in several critical technologies."

Trivelpiece Panel Report, January, 2010

Exascale Report: National Security

Science of Nonproliferation
(similar to biosurveillance requirements)

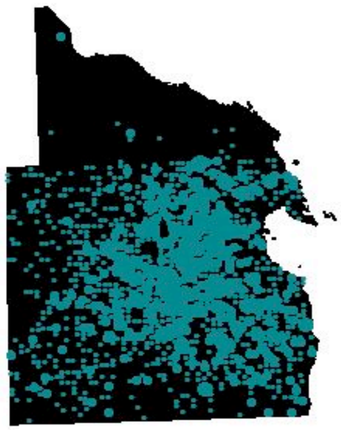
- Gather input
  - Build, design and interpret data from sensors
  - Uncertainty quantification
- Model problem
  - Proliferation process simulation
- Aggregate simulation results and observations
  - Data integration
- **Analyze results**
  - **Information exploration and analysis**
    - **Analyst in the loop**
    - Automated
      - Statistics and machine learning for detecting rare and anomalous behavior

Bovine Tuberculosis

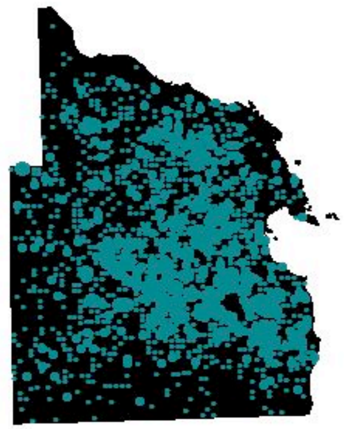
- Spread through the exchange of respiratory secretions.
- Minimal biosecurity at farms, having high density of cattle. Deer can wander in.
- TB can survive on feed for many days in a range of temperatures.
- Hunt clubs can create conditions leading to high density of deer.
- Lab experimentation expensive, requiring Biosafety 3 level labs suitable for wildlife.
- Because of this, details on species-specific susceptibilities not well understood.

Required hunting reports

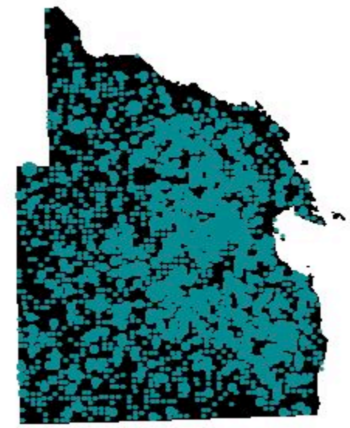
Deer kill location accurate to 1 square mile



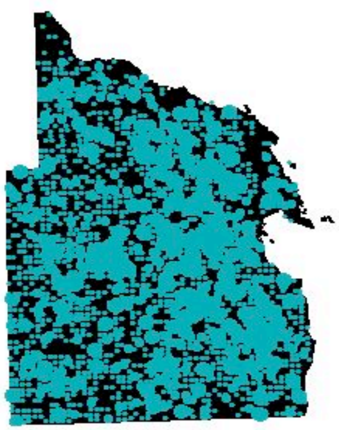
1996 Firearm Harvest



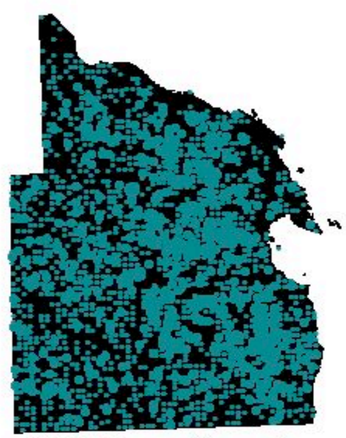
1997 Firearm Harvest



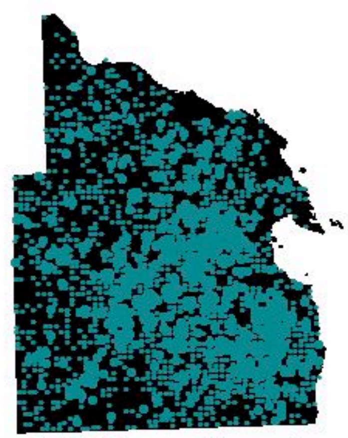
1998 Firearm Harvest



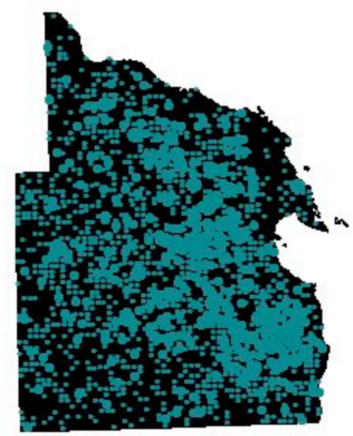
1999 Firearm Harvest



2000 Firearm Harvest



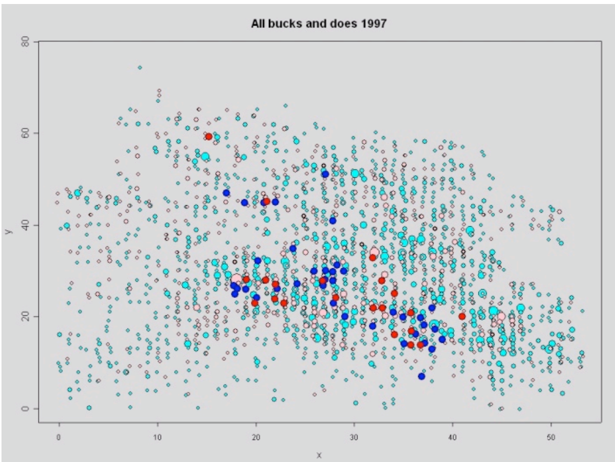
2001 Firearm Harvest

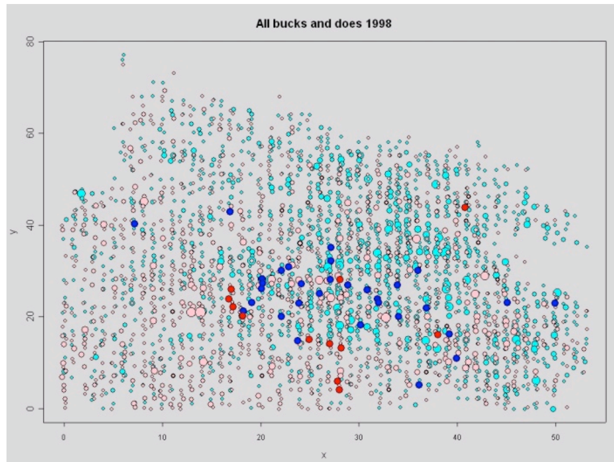


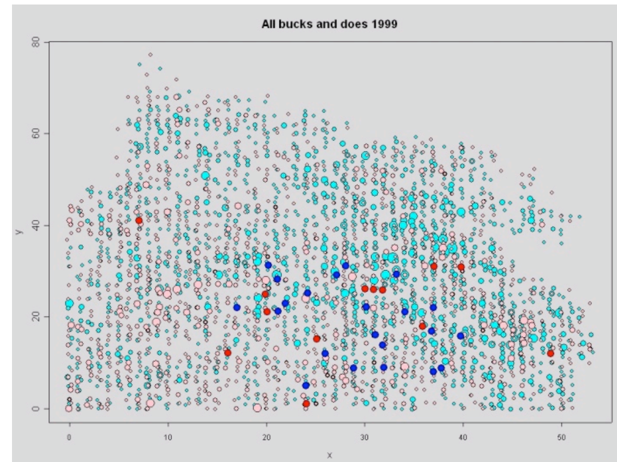
2002 Firearm Harvest

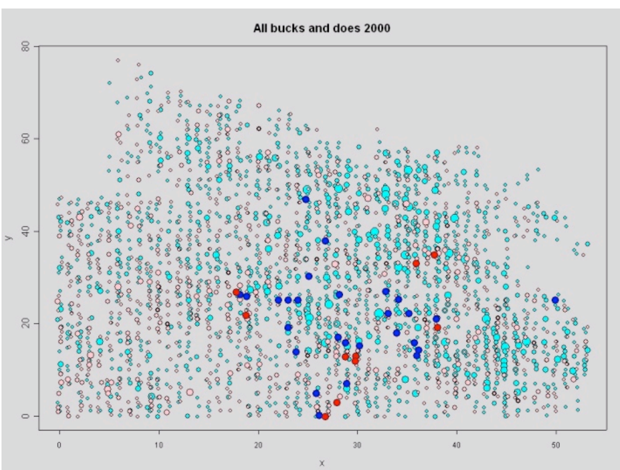
Tests on harvested deer yields prevalence maps of TB

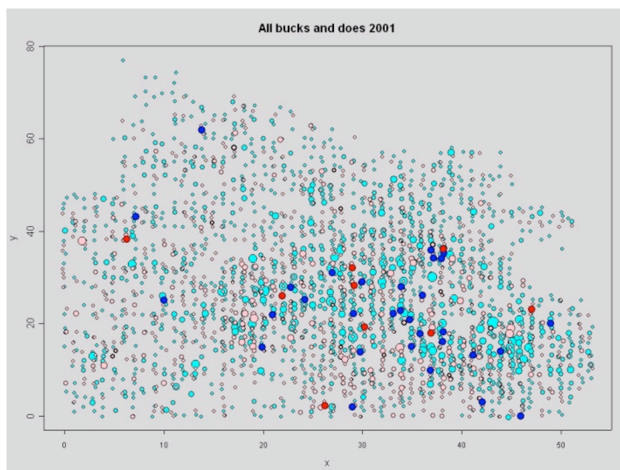
TB+ buck
TB+ doe
TB- buck
TB- doe











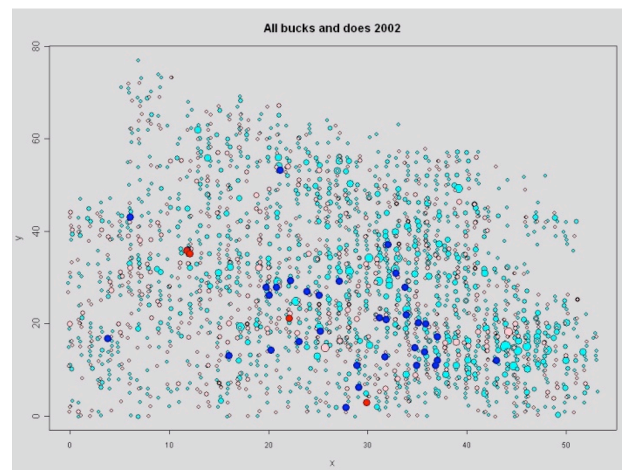
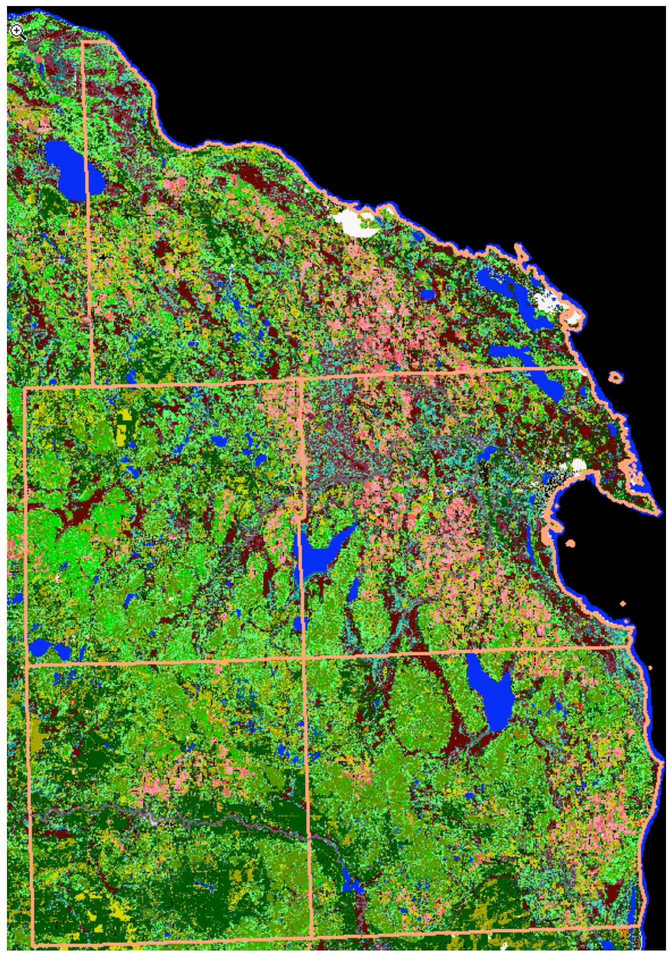


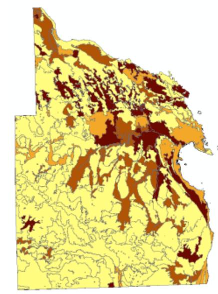
Image processing from satellite data yields a deer's view of the landscape

USGS 30 meter land-use codes

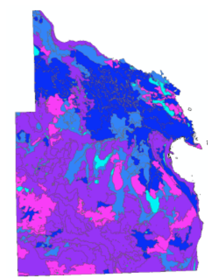
Deer habitat shape data



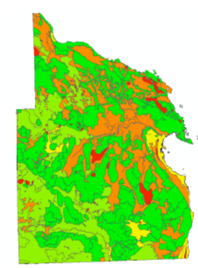
Therma
l
Cover ( ) =



Fall/Winter Habitat ( ) =

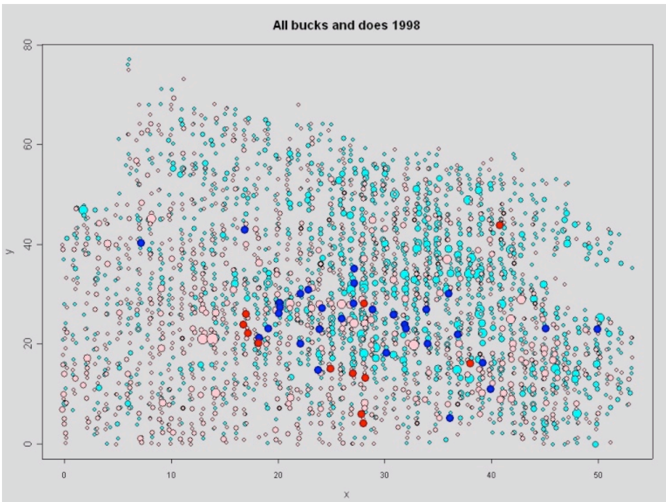


Spring/Summe
r
Habitat ( ) =



Populate the agent simulation

Harvest density data
(location, age, sex, TB +/-)

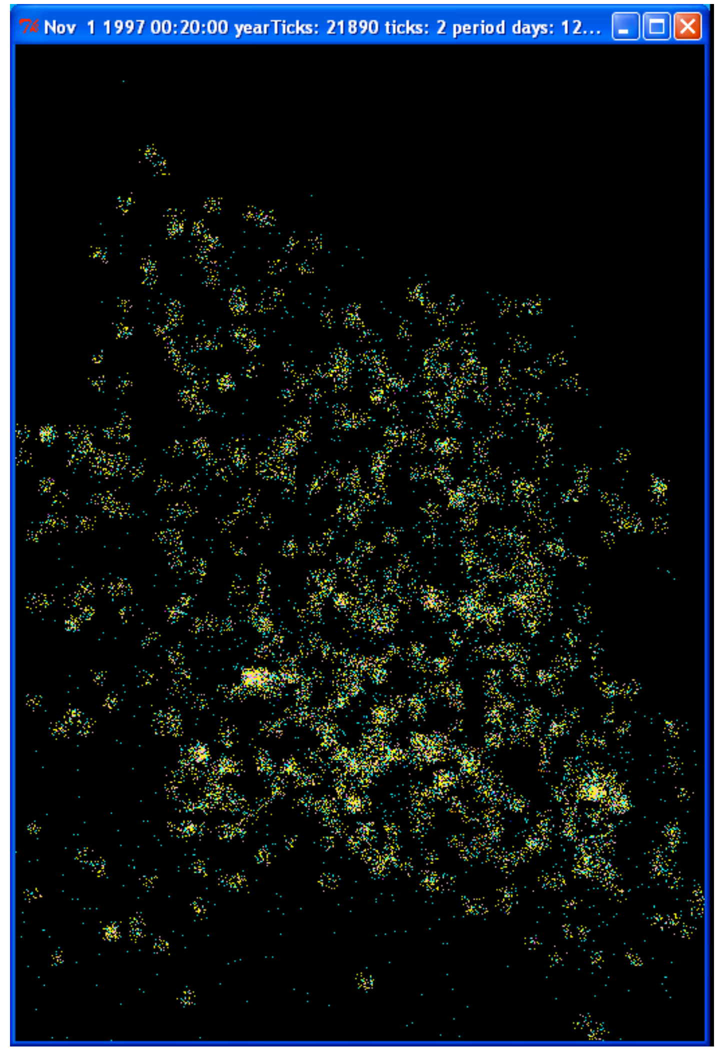


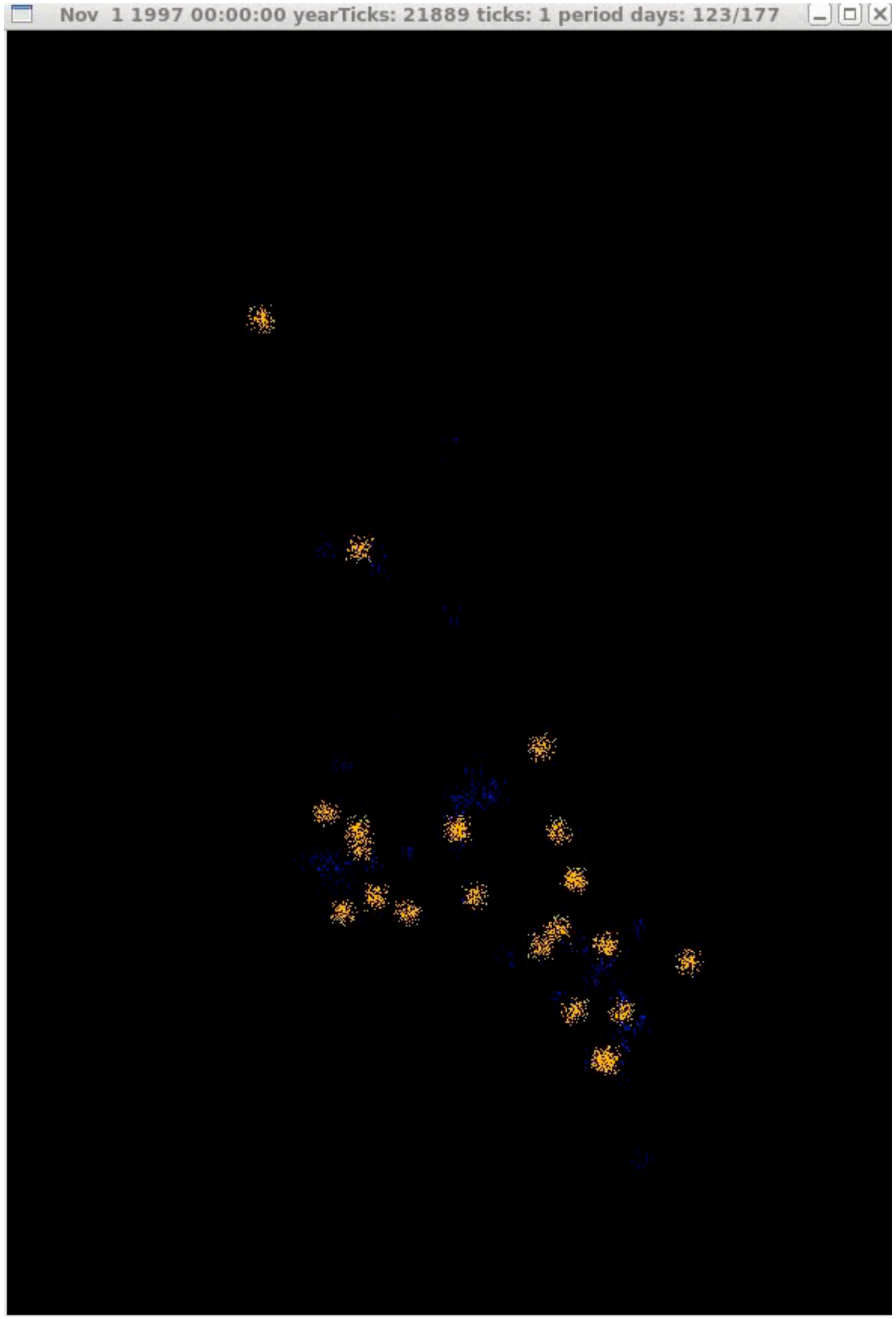
+

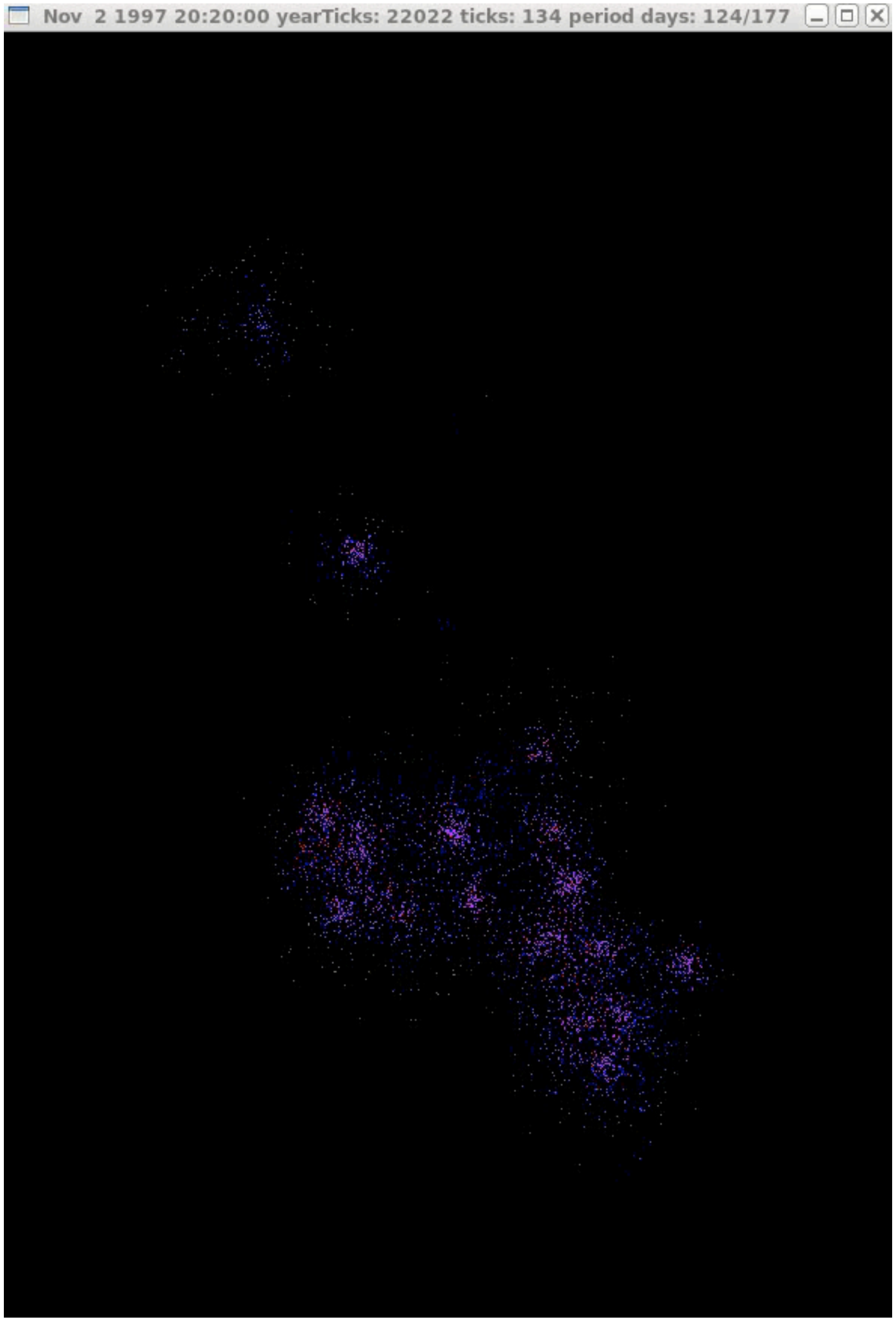
	Yearling Buck	Yearling Doe	Adult Buck	Adult Doe
1997	10	171	10	100
1998	11	180	14	120
1999	12	150	12	110
2000	13	140	20	140
2001	12	120	8	150
2002	11	110	10	160

=

Spatially explicit estimate of true popula




Nov 1 1997 00:00:00 yearTicks: 21889 ticks: 1 period days: 123/177


Nov 2 1997 20:20:00 yearTicks: 22022 ticks: 134 period days: 124/177

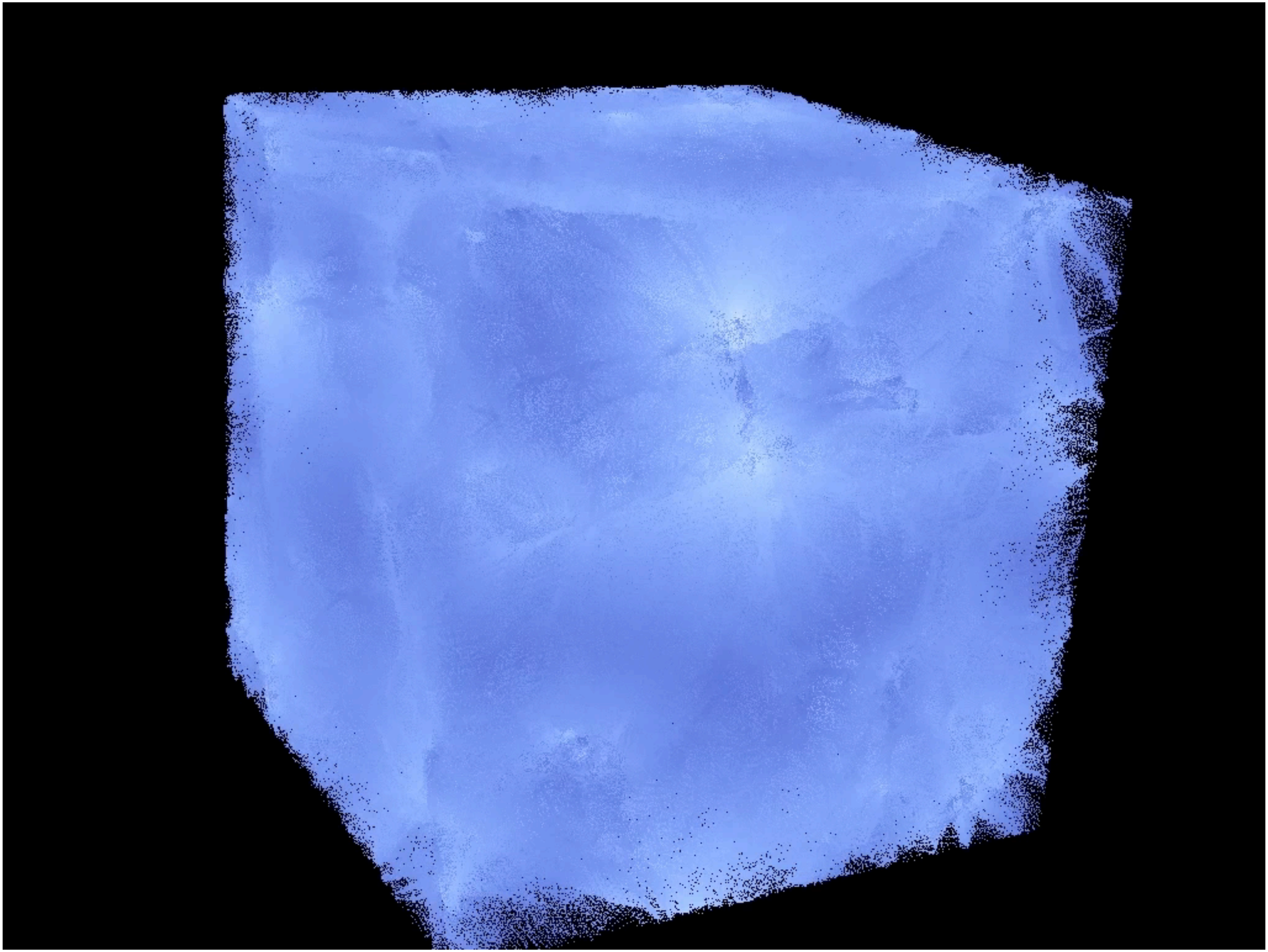
Exascale data analysis capability

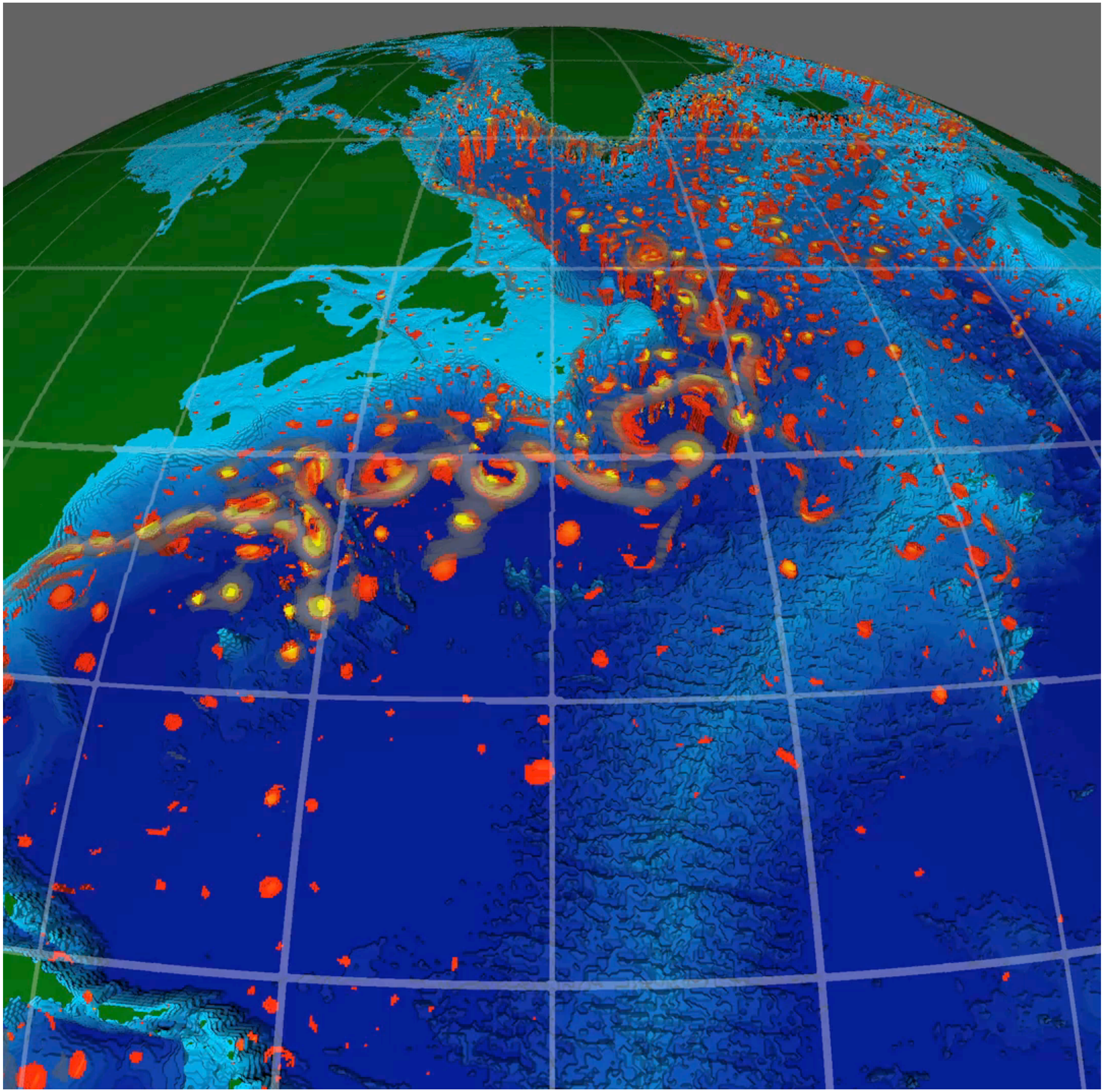
Prefix	Mega	Giga	Tera	Peta	Exa
10^n	10^6	10^9	10^{12}	10^{15}	10^{18}
Technology	Displays, networks		Data sizes and machines		

A "Middle Way" between:

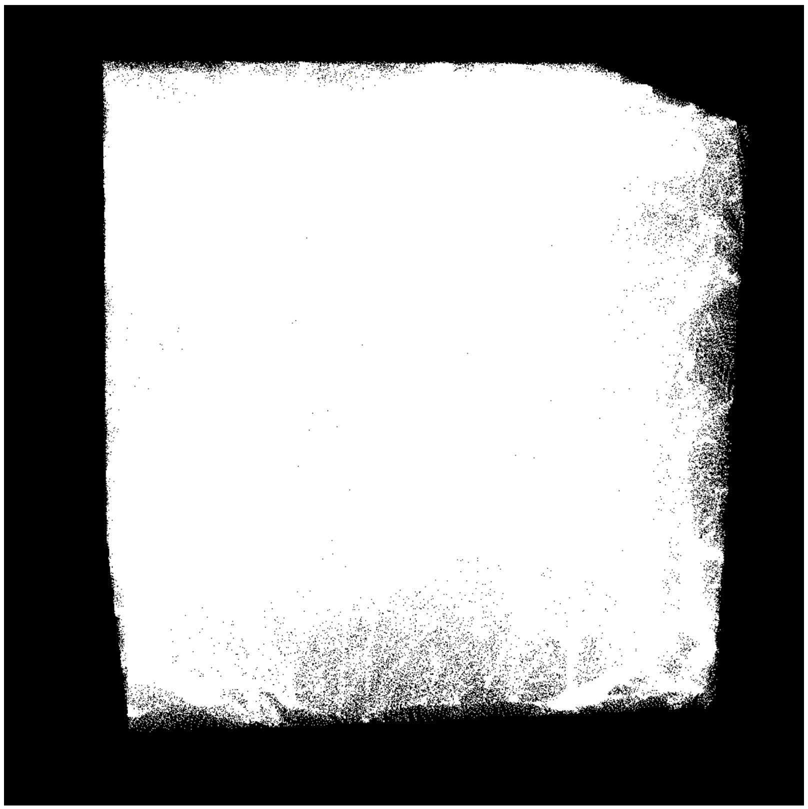
- Numerically-intensive / HPC approach
  - Massive FLOPS
    - Top 500 list – 1999 Terascale, 2009 Petascale, 2019? Exascale
    - Roadrunner – First petaflop supercomputer – Opteron, Cell
- Data-intensive supercomputing (DISC) approach
  - Massive data

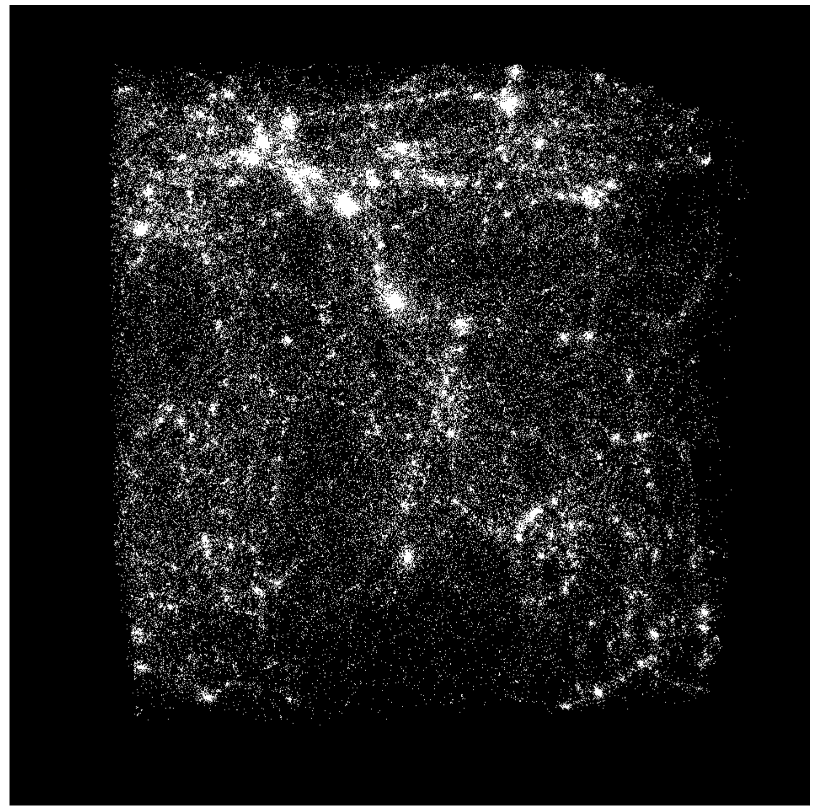
- We are exploring it by necessity for interactive scientific visualization of massive data
  - DISC using a traditional HPC platform





Multi-Resolution Sampling



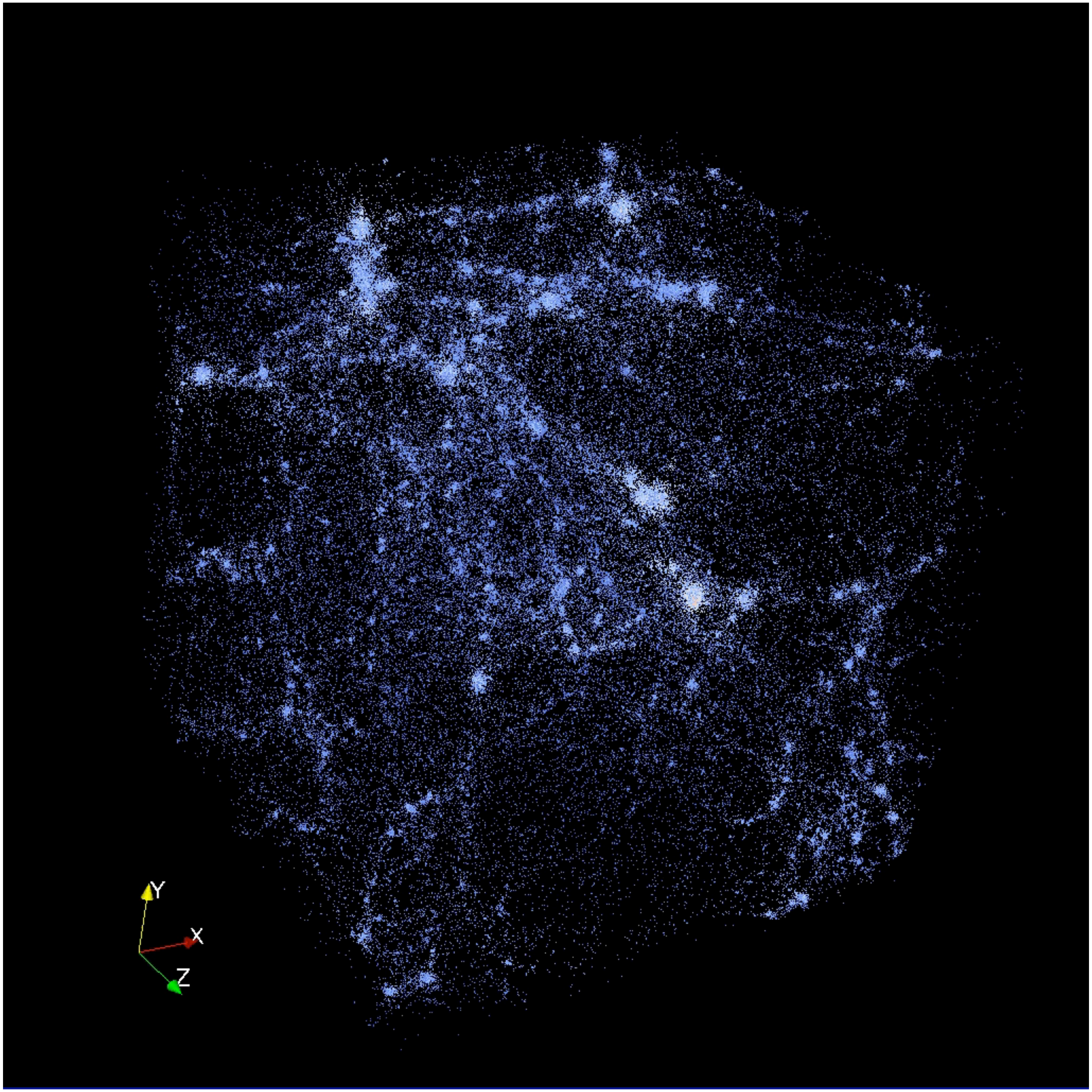


Sampling

- In-situ & storage-based sampling-based data reduction
  - Can work with all data types (structured, unstructured, particle) and most algorithms with little modification
- Intelligent sampling designs to provide more information in less data
  - Little or no processing with simpler sampling strategies (e.g., pure random)
- Untransformed data with error bounds
  - Data in the raw; Ease concerns on unknown transformations/alterations
  - Probabilistic data source as a first-class citizen in visualization and analysis

Sampling

- Quantify the data error for analysis, quantify visual error for vis
  - Show the data error, allow the user to reduce error incrementally
  - Scientist is always informed of the error in their current view
- Data size scales with sample size for bottlenecks
  - Any sample sizes based on error constraints and system/human constraints
  - Same model could be used in simulations to reduce data output per time step


Y
X
Z

Conclusions

- There are many opportunities for supercomputing in the field of biosurveillance

Acknowledgements

- Team: John Patchett Jonathan Woodring Li-Ta Lo Susan Mniszewski Patricia Fasel Joshua Wu Christopher Mitchell Sean Williams

- Support:
- Los Alamos National Laboratory – LDRD
  - Cosmology and sampling project
- USDA
  - Bovine tuberculosis Project
- DOE Office of Science
  - Climate modeling